PATENT
Docket No. CH9-2000-0004(246)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application Dieter Jaepel *et al.*

Application No. Examiner:

Filed: (Herewith) Group Art Unit:

For: IMPROVED SPEECH RECOGNITION BY AUTOMATED CONTEXT CREATION

CLAIM OF FOREIGN PRIORITY

Box Patent Application
Commissioner for Patents
Washington, D.C. 20231

Sir:

Priority under the International Convention for the Protection of Industrial Property and under 35 U.S.C. §119 is hereby claimed for the above-identified patent application, based upon European Application No. 00116450.8, filed July 28, 2000, and a certified copy of this application is submitted herewith which perfects the Claim of Foreign Priority.

Respectfully submitted,

Date: 7.20.01

Gregory A. Nelson, Reg. No. 30,577
Kevin T. Cuenot, Reg. No. 46,283
Steven M. Greenberg, Reg. No. 44,725
AKERMAN SENTERFITT
222 Lakeview Avenue, Suite 400
Post Office Box 3188
West Palm Beach, FL 33402-3188
Telephone: (561) 653-5000

Express Mail Label No. EL920516495US

P1017717;1

CH9-2000-0004

Europäisches Patentamt	European Patent Office	Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.	The attached documents are exact copies of the European patent application described on the following page, as originally filed.	Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

00116450.8

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 28/06/01
LA HAYE, LE

Europäisches Patentamt | European Patent Office | Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 00116450.8
Demande n°:

Anmeldetag:
Date of filing: 28/07/00
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
International Business Machines Corporation
Armonk, NY 10504
UNITED STATES OF AMERICA

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Improved speech recognition by automated context creation

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

IMPROVED SPEECH RECOGNITION BY AUTOMATED CONTEXT CREATION

FIELD OF THE INVENTION

The present invention relates to the field of speech processing and speech recognition in general. In particular, the invention relates to systems and methods for generating an output by means of a speech input.

BACKGROUND OF THE INVENTION

Due to the recent advances in computer technology as well as the recent advances in the development of algorithms for the processing and recognition of speech, speech recognition systems have begun to become increasingly more powerful and increasingly less expensive.

Certain speech recognition systems match the words to be recognized with words of a vocabulary. The words in the vocabulary are usually represented by word models (word baseforms). A word can for example be represented by a sequence of Markov models. The word models are used in connection with the speech input in order to match the input to the words in the vocabulary.

Most of today's speech recognition systems are constantly improved by providing larger vocabularies or by increasing the recognition rate by employing better algorithms. Such systems typically have 100'000 words and some products, such as IBM's ViaVoice software, have even 240'000 word entries.

Many of the commercially available speech recognition systems operate by comparing a spoken utterance against each word in its vocabulary. Since each such comparison can require thousands of computer instructions, the amount of computation required to recognize an utterance grows dramatically with increasing vocabulary size. This increase in computation has been a major problem in the development of large vocabulary systems.

Some speech recognition systems can be trained by the user uttering a training text of known words. This allows to tailor the system to a particular user. This training leads to an improved recognition rate.

There are bi-gram and tri-gram based recognition systems that search for like-sounding words such as 'to', 'two', and 'too', by analyzing them either in a context of two consecutive words (di-gram technology) or three consecutive words (tri-gram technology). The di-gram technology and the tri-gram technology leads to an improved recognition rate.

It is a problem of such systems, that as the vocabulary grows, the number of words that are similar in sound also tend to grow. As a result, there is an increased likelihood that an utterance of a given word from the vocabulary will be mis-recognized as corresponding to another similar sounding word from the vocabulary.

There are different approaches known for reducing the likelihood of word confusion. One such method is called "pruning". Pruning is a common computer technique used to reduce the computation. Generally speaking, pruning reduces the number of cases which are considered by eliminating some cases from further consideration. Scores (representing the likelihood of occurrence in an input) can be assigned to the words in a vocabulary in order to be able to eliminate words from consideration during the recognition task. The score is updated during the recognition task and words that are deemed irrelevant for the recognition are not considered anymore.

Another technique to cope with large vocabulary systems is that of hypothesis and test, which is, in effect, a type of pruning, too. When features are observed in a speech input, they are used to form a hypothesis that the word actually spoken corresponds to a subset of words from the original vocabulary. Then the speech input is processed further by performing a more lengthy match of each word in this sub-vocabulary against the acoustic signal received. This sub-vocabulary is directly derived from the speech input

Yet another approach for dealing with the large computational demands of speech recognition in large vocabulary systems, is the development of special purpose hardware to increase greatly the speed of such processing. There are for example special purpose processors that perform probabilistic frame matching at high speed.

There are a host of other problems which have been encountered in known speech recognition systems, such as background noise, speaker dependent utterance of words, insufficient processing speed, just to mention some of the problems. All these disadvantages and problems have so far prevented a widespread use of speech recognition in many market domains.

Despite the recent advances in speech recognition, there is a great need to improve further the performance of speech recognition systems before they will find an even larger distribution in the market.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a speech processing system and method which increases the ease of use.

The method according to an illustrative embodiment of the present invention proposes a procedure where a voice-generated output is generated using a computer system. The output is generated by receiving an input and automatically creating a context-enhanced database using information derived from the input. The voice-generated output is generated from a speech signal by performing a speech recognition task to convert the speech signal into computer processable segments. During this speech recognition task the context-enhanced database is accessed in order to improve the speech recognition rate, i.e., to interpret the speech signal in light of words given in the context-enhanced database.
The user is enabled to edit/correct the output in order to generate a final output which is then made available.

A speech processing system, in accordance with the present invention, is able to generate a voice-generated output. The system comprises a module for automatically creating a context-enhanced database by using information derived from a system input, and a speech recognition system for converting a speech signal into segments. The context-enhanced database is accessed in order to find matching segments. The system further comprises a module for preparing the voice-generated output with the matching segments, and a module for enabling editing/correction of the output to generate a final output which is then made available.

According to the present invention, the number of words which undergo an extensive match (e.g., an acoustic match) against uttered words is drastically reduced.

The present scheme allows to realize implementations that are less expensive and computationally less demanding. In other words, the present invention can even be used in smaller systems that are not as powerful as today's desktop computers.

Advantages of the present invention are addressed in connection with the detailed description or are apparent from the description.

DESCRIPTION OF THE DRAWINGS

The invention is described in detail below with reference to the following schematic drawings.

FIG. 1 shows a schematic block diagram of a conventional speech recognition system.

FIG. 2 shows a schematic block diagram of a first speech processing system, according to the present invention.

FIG. 3 shows a schematic block diagram of a second speech processing system, according to the present invention.

FIG. 4 shows a schematic block diagram of a building block of the first speech processing system, according to the present invention.

FIG. 5 shows a schematic block diagram of a third speech processing system, according to the present invention.

FIG. 6 shows a schematic block diagram of a fourth speech processing system, according to the present invention.

FIG. 7 shows a schematic block diagram of a fifth speech processing system, according to the present invention.

FIG. 8 shows a screen shot of a window as produced by a speech processing system, according to the present invention.

DETAILED DESCRIPTION

According to the present invention, a scheme is provided that greatly simplifies the interaction between the user and a computer system. It is herein proposed to use input information that is available in order to provide for a better and more reliable speech recognition.

If a user works with a computer system, there is usually at least one active application program, i.e., a program that is currently used by the user. It is assumed that the user is working on/with this active application program. The active application program is in many cases closely related to the user's current work task.

This can be illustrated by means of a simple example. Assuming that the user of a computer system (recipient) has received an electronic mail (E-mail) from another user, it is likely that the recipient will open the E-mail in order to print or read it. It is further assumed that the other user is expecting the recipient to respond to this E-mail. This means that the respective mailer software (e.g., Lotus Notes) is active and that the E-mail is displayed in a window on the computer's screen. It is highly likely, that the contents of this E-mail defines the context for the recipient's response. Input information can thus be derived from this E-mail.

According to the present invention, input information is derived in a pre-processing step which defines the contents for an output that is to be generated by the user of the computer system. In the above example, the input information can be derived from the text contained in the E-mail received. It is, however, also possible

1. to derive the input information from the history of E-mails (e.g., a chain of inquiries and responses to these inquiries),
2. to derive the input information from a document that is currently on the computer screen (e.g., a scientific paper currently read by the user),
3. to derive the input information from a chain of related documents,
4. to derive the input information from linked documents,
5. to derive the input information from a specific folder or directory,
6. to derive the input information from the attachments that are received with an E-mail,
7. to derive the input information from a spread sheet currently used by the user,
8. to derive the input information from the computer's cache memory,
9. to derive the input information from the history information recorded by a web browser,
10. to derive the input information from a knowledge management system,
11. to derive the input information from an incoming message, e.g., an incoming request in a call center,
12. to derive the input information from a received facsimile,
13. to derive the input information from the result of a database search, and so forth.

This input information, no matter how it is generated, is assumed to define the context in light of

which the user is expected to generate an output, as mentioned above. According to the present invention, the user is enabled to generate this output by uttering words. The respective output is thus referred to as voice-generated output.

The voice-generated output can be an E-mail, a facsimile, a letter, a memo or any other output (e.g., a reaction) that can be generated by a computer system.

In order to be able to prepare the voice-generated output, the user is requested to utter words. This speech input undergoes a speech recognition task after having been transformed from an audio signal into a signal format that can be processed by a computer system.

For this purpose, an audio system is employed. The audio system may comprise a microphone, a microphone followed by some audio processing unit(s), or similar means. The audio system is employed to receive the words uttered by the user of the speech recognition system, to transform the uttered words into an audio signal, and to feed this audio signal to a converter system.

The converter system may comprise an analog-to-digital (A/D) circuit, for example. It is the purpose of the converter system to convert the audio signal into a signal format that can be processed by the computer system. In most implementations of the speech recognition system according to the present invention, the converter system generates a digital signal.

According to the present invention, a speech recognition task is performed in order to convert the uttered words into computer processable segments, such as word segments (e.g., letters or syllables), phonemes, phonetic baseforms, frames, nodes, frequency spectra, baseforms, word templates, words, partial sentences, and so forth.

Computer processable in the present context means a representation that can be processed by a computer system.

In order to be able to do this in an efficient and reliable manner, a context-enhanced database is

generated using the input information received. The context-enhanced database can either be directly derived from the input information, or it can be derived from an existing database using the input information. The input information can be used for example, to define a smaller, specific portion within a pre-installed large lexicon. A context-enhanced database may comprise between a few words up to a few thousands words, preferably between 10 words and 1000 words. The size of the context-enhanced database depends on the actual implementation of the inventive scheme and on the size of the input itself. According to the present invention, the context-enhanced database is dynamically generated (updated) depending on or taking into account the user's current or most recent activities.

As mentioned in the last section, the context-enhanced database can either be generated directly from the input information, or it can be derived from an existing database using the input information. The latter can be done by generating a word list from the input information (e.g., by extracting words from an E-mail to be responded to) and by connecting/linking this word list to an existing lexicon. The word list can be connected/linked to the lexicon such that it acts as a filter or first instance that is accessed during a speech recognition task. The underlying lexicon is only accessed if no matching word was found in the word list. There are others ways of implementing this aspect of the invention, as will be addressed later-on.

During the speech recognition task, the context-enhanced database is accessed in order to improve the speech recognition rate. The segments derived from the words uttered by the user when preparing an output are interpreted in light of the words given in the context-enhanced database. According to the present invention, the number of processable segments which undergo an extensive match (e.g., an acoustic match) against uttered segments is drastically reduced, since the matching is done – at least in a first run – with information in the context-enhanced database only.

According to the present scheme, the output is prepared while the user talks into the audio system. In a subsequent step, the system might enable the user to edit or correct the output in order to generate a final output. There are different approaches that can be used in order to

enable a user to edit or correct the output. The system may, for example, display the output on a screen to allow the user to read it and to intervene manually, if there is something to be edited or corrected. It is also conceivable that the system highlights those words where there is a certain likelihood of misinterpretation (mis-recognition) of the user's speech (unknown words, similar sounding words, etc). Other implementation examples are given in connection with specific embodiments.

After having finished the speech recognition task, the final output is made available for further processing. The final output can be sent via a mailer to another user, it can be prepared for printing, it can be mailed via a fax modem or a fax machine, it can be stored in a memory, and so on. For this purpose, the output is temporarily put into a memory from where it can be printed, transmitted, fetched by some other application program, or the like.

The present scheme improves known speech recognition schemes by providing a context-enhanced database which is derived from some input information that is assumed to be related to the user's current task. Due to this, the speech recognition can be performed in light of a well defined context rather than a huge lexicon. An output is generated by transcribing/synthesizing the human dictation in light of the a context-enhanced database.

The expression „computer system" is herein used as a synonym for any system that has some computational capabilities. Examples are: personal computers (PCs), notebook computers, laptop computers, personal digital assistants (PDAs), cellular phones, etc.

A speech recognition system is a system that performs a speech recognition task. Typically, a speech recognition system is a combination of a general purpose computer system with a speech recognition software. A speech recognition system can also be a special purpose computer system, such as a system with special purpose speech recognition hardware.

Speech recognition systems are marketed which run on a commercial PC and which require little extra hardware except for an inexpensive audio system (e.g., a microphone and an audio card

with an analog-to-digital (A/D) converter, and a relatively inexpensive microprocessor to perform simple signal processing tasks). Such systems are able to provide discrete word recognition. There are also computer system which just require a speech recognition software. The necessary hardware components are already present in form of an integrated microphone and an A/D converter.

A schematic representation of a conventional speech recognition system 10 is given in Figure 1. Most speech recognition systems 10 operate by matching an acoustic description of words (e.g. Word 14) in their lexicon 13 against a representation of the acoustic signal generated by the utterance of the word (e.g. Word' 11 received as an input 12) to be recognized. If the input word 11 matches the word 14 in the lexicon 13, then an output 15 is generated comprising the word 14. The speech signal representing the word 11 is converted by an A/D converter into a digital (signal) representation of the successive amplitudes of the audio signal created by the speech. Then that digital signal is converted into a frequency domain signal which consists of a sequence of frames, each of which gives the amplitude of the speech signal in each of a plurality of frequency bands. Such systems commonly operate by comparing the sequences of frames produced by the utterance to be recognized with a sequence of nodes, or frame models, contained in the acoustic model of each word in their lexicon. Such a speech recognition system is called a frame matching system.

The performance of frame matching systems can be improved using a probabilistic matching scheme and a dynamic programming scheme, both of which are known in the art for some time now. The application of dynamic programming to speech recognition is described in the article "Speech Recognition by Machine: A Review" D.R. Reddy, in Readings in Speech Recognition, A. Waibel and K.-F. Lee, Editors, 1990, Morgan Kaufmann: San Mateo, CA, pp. 8 – 38.

One embodiment of a speech processing system 20, according to the present invention, is illustrated in Figure 2. In a pre-processing step, a context-enhanced database 21 is generated from input information 22, as described in one of the previous sections. If now a speech signal is received by an audio system 24, as indicated by arrow 23, an audio signal representing this

speech signal is forwarded via line 25 to a converter system 26. The audio system 24 transforms the acoustic signal received via 23 into the audio signal. This audio signal is fed via line 25 to the converter system 26 where it is transformed into a signal format that is processable by the speech recognition engine 27. In most implementations, the converter system 26 is designed to generate a digital signal that is fed via line 28 to the speech recognition engine 27. This digital signal represents processable segments uttered by a user. The speech recognition engine 27 now matches the processable segments with segments in the context-enhanced database 21, as indicated by the arrow 29. All those segments for which a matching segment was found in the context-enhanced database 21 (called matching segments) are fed to an output unit 30 where an output is generated. The user now can interact with the system 20 by editing and/or correcting the output, as indicated by the output editing/correction unit 31. The user interaction is illustrated by the arrow 32. The unit 31 provides a final output 33 at an output line 34. Depending on the implementation, some of the steps can be performed concurrently.

Another embodiment of a speech processing system 40, according to the present invention, is illustrated in Figure 3. In this example, an E-mail 41 from an E-mail folder 42 delivers the input information. The E-mail folder 42 may be the inbox of a mailer 43. The mailer 43 also has an outbox 44. There is one E-mail 45 waiting in the outbox 44 for delivery. As schematically shown in Figure 3, the E-mail 41 contains the usual address information 47, a subject field 46, and a text body 48. According to the present embodiment, a word list 49 is derived from the contents of the E-mail 41. The word list 49 can be derived from the address information 47, the subject field 46, the text body 48, or from a combination thereof. This word list 49 is used to provide a context-enhanced database (not shown). In the present embodiment, the word list 49 sits on top of a lexicon 13 that has many word entries.

If the user now wants to prepare an output (e.g., a response to the E-mail 41), he activates the speech recognition module and talks into a microphone, for example. The respective speech signal (box 50) is analyzed by a conventional phoneme processing engine 51. Then a word matching process is carried out by the word matching engine 52. This word matching engine 52 has an application programming interface (API) 53 that serves as an interface for accessing a lexicon. A conventional speech recognition system would access a large lexicon through this

interface 53 in order to find matching words. According to the present invention, the word list 49 is accessed first through the API interface 53. If all words uttered by the user and represented by the speech signal are found in the word list 49, a grammar check may be performed by a grammar check unit 54 before an output 57 is generated by the output generation unit 55. This output 57 is provided at the output line 56 for further processing. In the present embodiment, the output 57 is the body of an E-mail that is stored in a memory unit 58. It can be fetched from this memory 58 and pasted into an outgoing E-mail. The E-mail 45 that sits in the outbox of the mailer 43 was generated exactly the same way. As soon as the computer system 40 connects to a network, the outgoing mail can be transmitted.

The word matching engine 52 may be implemented such that is always returns the best match for a word received from the unit 51. Part of the output may be presented to the user right away at output line 56 before he has completed spelling the desired words.

Advantageously, the speech processing system 40 may be implemented in such a way that the lexicon 13 is accessed if there are words for which no matching counterpart was found in the word list 49. This can be done through the same API interface 53, or a separate API interface may be provided for that purpose.

The pre-processing module 36 that performs the pre-processing steps described in connection with the system implementation illustrated in Figure 2, is schematically summarized in Figure 4. As illustrated in this Figure, some input information 22 is received via an input line 35. This input information 22 may stem from an E-mail currently processed in an editor or from a some other source, as indicated by the items 1. – 13. in the above listing. The context-enhanced database 21 is automatically created by deriving information from the input information 22. There is some kind of an interface 29 that allows the speech recognition engine 27 (cf. Figure 2) to access the context-enhanced database 21. This context-enhanced database 21 then returns matching segments or matching words.

Another pre-processing module 65 is shown in Figure 5. The input information 22 is received

via an input line 63. A processing unit 60 is employed, that takes information (e.g., segments or words) from the input information 22 and creates a context-enhanced database 62. In order to obtain a better context-enhanced database 62, a synonym lexicon 61 is employed. If the input information comprises a WordA, the processing unit 60 creates several entries in the context-enhanced database 62; one for the original word, namely WordA, and as many entries as there are synonyms in the synonym lexicon 61. Assuming that there are three synonyms WordA', WordA'', and WordA''' in the lexicon 61, four entries (WordA, WordA', WordA'', and WordA''') will be created in the context-enhanced database 62. If the user of a system, in accordance with the present invention, utters a word that is a synonym to a word comprised in the input information 22, the system will be able to recognize this word and add it to the output currently generated. There is an interface 64 (e.g. a standardized interface) that allows a speech processing system to access the context-enhanced database 62.

Yet another pre-processing module 75 is depicted in Figure 6. In this example, the input information 22 is received via an input line 73. A processing unit 70 is employed, that takes information (e.g., segments or words) from the input information 22 and builds a context-enhanced database 72. In order to obtain a context-enhanced database 72 with more word entries, a database 76 with meaning variants and a synonym lexicon 77 are employed. If the input information comprises a WordA (e.g., the word "plant"), the processing unit 70 will access the meaning variants database 76 in order to check whether there is more than one meaning for the WordA. In case of the word "plant", the database 76 will comprise two entries. The first entry (WordA*) identifies the "living plant" and the second entry (WordA**) identifies the "building" or "industrial fabrication plant". Both these meaning variants (WordA* and WordA**) are retrieved by the processing unit 60. Other information can be used by the processing unit 60 to identify which of the two variants (WordA* or WordA**) is the one that is actually meant. If the input information contains the sentence "A plant was erected in 1985", for example, then it is clear from the context that the building (WordA**) and not the living object (WordA*) is referred to. The synonym lexicon 77 now delivers synonyms for this second variant (WordA**). This scheme allows the system to avoid misunderstandings due to the misinterpretation of different word variants. It resolves these issues while creating the

context-enhanced database 72. The context-enhanced database 72 is accessible via the interface 74.

The embodiment illustrated in Figure 7 is even more powerful than the previous ones, since it employs a meaning extraction system 81 in connection with a knowledge database 86. The module 85 comprises a memory with the input information 22. The processing unit 80 consults the meaning extraction system in order to get some understanding of what is contained in the input information 22. The meaning extraction system can be a system interacting with a fractal hierarchical knowledge database 86, as for example described and claimed in the European patent application entitled "Processing of textual information and automated apprehension of information", filed on 2 June 1998, currently assigned to the assignee of the instant patent application. Such a meaning extraction system 81 is able to understand – at least to some extent – what is meant by the input information 22. It is even able to extract additional information that is believed to be associated or related. The processing unit 80 is thus able to build a context-enhanced database 82 that is 'richer' in that it not only contains the words that were found in the input information, but also information that is deemed to be related. The context-enhanced database 82 is accessible via the interface 84 and the input information 22 is received via an input line 83.

An example of a speech recognition system is illustrated in Figure 8. The front-of-screen of a simple recognition system is shown. It comprises an editor 90 with a text-editing window 91. The speech recognition system displays the result of a speech recognition exercise where the user uttered the partial sentence "...plant causes pollution ...".
The recognition worked fine, except for the word "plant" which was misunderstood by the system. As shown in the text-editing window 91, the system transforms the partial sentence and displays the resulting output. 92. The system outputs the word "plan" instead of "plant". The user now is able to edit/correct the output 92. In order to do so, he can double-click on a word that is believed to be misspelled and the system highlights the respective word. In the present example, the word "plan" is highlighted using the mouse. A correction window 93 is opened that now offers different alternatives 1 – 2 for the word "plan". If a meaning extraction system, like

the one in Figure 7, is employed in the background, the processing system is able to tell from the context of the context-enhanced database that an industrial fabrication plant is meant. The system thus is able to offer the best matching word in the uppermost position 1 in the correction window 93. The system is able to determine, that the other word "planned" is not likely to be relevant since this word does not make sense in the present context. By clicking on the OK-button, the word "plan" can be corrected so that it reads "plant".

A speech recognition system according to the present invention can be realized such that the word "plan" is automatically corrected. This is possible, since the system is able to recognize that the word "plant" is the only word that makes sense in the present context.
An implementation of the present invention that makes use of a word list (context-enhanced database) generated from an active window (e.g., an E-mail) would be able to check whether the word "plan" is comprised in the context-enhanced database. If this word is not in the context-enhanced database, then the system would replace it by the word "plant", provided that the word "plant" is in the context-enhanced database. A system according to the one illustrated in Figure 7 would be able to determine that the combination of the words "pollution" and "plant" is valid and that the combination of "pollution" and "plan" is not valid. This also allows an automatic correction.

According to one embodiment of the present invention, a template (form) is automatically generated from the input information. The voice-generated output can be inserted into the template. Such a template-based approach is well suited for situations where a highly automated response is required and where the responses typically all look the same. An example could be a booking system used by a chain of affiliated hotels.

The present invention can be used in connection with systems that process discrete speech (e.g., word-by-word) or continuous speech.

Advantageously, a system according to the present invention may comprise a speech synthesizer that converts the final output into a speech output. Such a speech synthesizer may comprise

synthesizer hardware with a parameter store containing representations of words to be output, and a loudspeaker, for example.

It is advantageous to provide an implementation of the present invention where a fall-back mode or procedure is realized that kicks in those situations where no matching words are being found. Such a fall-back mode or procedure may offer the user a simple interface for typing the missing words.

According to another embodiment of the present invention, the context-enhanced database is dynamically generated while input information is received. A first guess context-enhanced database is generated and then constantly updated as additional input information is received. An example is used to better illustrate this. A call is received on a call-in line of a call center. The call center system routes the call to an automated call handler which asks questions. The caller either responds by uttering words or alternatively by pressing buttons on the phone. While this goes on for a while, a first guess of a context-enhanced database can be generated. If the caller is not calling for the first time, caller specific information can be fetched from a memory. This caller specific information may be used to generate a context-enhanced database, or an old context-enhanced database may be retrieved that was generated during a previous call of the same caller. The context-enhanced database is now constantly updated as the caller reveals additional information about the reason for calling. An output may now be generated (e.g. a confirmation fax) by the operator of the system. In order to do so, the operator talks into a microphone. The words he utters are transformed and processed referring to the most current version of the context-enhanced database. The final output is temporarily stored, printed, signed and faxed to the caller's fax number.

With the present invention one is able to transcribe human dictation into an output, such as a letter or an E-mail. This greatly increases the speed and ease with which humans can communicate with other humans using computer-generated letters or E-mail. It makes it much easier for a human to record and/or organize their own words and thoughts. This can be done by storing a voice-generated output in a database, or by using the voice-generated output to update a

knowledge database.

It is another advantage of the present scheme that it can be used on PDA or phone-like systems which have no adequate keyboard.

With the proposed scheme, the speed of retrieval and the recognition rate can be improved since the context-enhanced database enables a faster and more reliable matching.

CLAIMS

1. Method for providing a voice-generated output (33) using a computer system, comprising the steps:

- receiving an input (22);
- automatically creating a context-enhanced database (21) by using information derived from the input (22);
- preparing output (30) from a speech signal (23) by performing a speech recognition task to convert the speech signal (23) into said output (30) comprising computer-processable segments , whereby the context-enhanced database (21) is accessed in order to improve the speech recognition rate;
- enabling editing of the output (30) in order to generate as final output (33) the voice-generated output (33); and
- making the final output (33) available.

2. The method of claim 1, whereby words are processed separately during the speech recognition task by

- identifying a matching word in the context-enhanced database (21) for each of the words, and
- adding the matching word to the output (30).

3. The method of claim 1 or 2, whereby during the speech recognition task the speech signals (23) are analyzed.

4. The method of claim 2, whereby another database (13) is accessed in order to find a matching word for each of the words for which no matching word was found in context-enhanced database (49).

5. The method of claim 1, whereby at least two of the steps of claim 1 are carried out concurrently.

6. The method of claim 1, whereby the processable segments are processable words.

7. The method of claim 1, whereby the speech signal (23) is interpreted as part of the speech recognition task in light of words given in the context-enhanced database (21).

8. The method of claim 1, whereby the input (22) is received from an application program..

9. The method of claim 1, whereby the input (22) is received from one or more of:

- a history of E-mails,
- a document that is currently on a screen of the computer system,
- a chain of related documents,
- linked documents,
- a folder or directory,
- attachments that are received with an E-mail,
- a spread sheet,
- a cache memory of the computer system,
- a history information recorded by a web browser,
- a knowledge management system,
- an incoming message,
- a received facsimile, and
- from a result of a database search.

10. The method of claim 1, whereby the context-enhanced database (21) defines a context in light of which the voice-generated output (33) is to be generated.

11. The method of claim 1, whereby the voice-generated output is a physical output.

12. The method of claim 11, whereby the voice-generated output is temporarily put into a memory (58).

13. The method of claim 1, whereby the editing/correction is enabled by highlighting those words of the output where there is a certain likelihood of misinterpretation (mis-recognition) of the speech signal (23).

14. The method of claim 1, whereby the context-enhanced database is derived from an existing database (61; 76, 77; 86) using the input.

15. The method of claim 1, whereby the context-enhanced database is dynamically generated and/or updated.

16. The method of claim 1, whereby one or more of a synonym lexicon (61) and a meaning variants database (76) is accessed when preparing the voice-generated output.

17. A speech processing system (40) for providing a voice-generated output (57), the system (49) comprising:

- a module for automatically creating a context-enhanced database (49) by using information derived from an input (41);
- a speech recognition system (51 – 56) for converting a speech signal (50) into segments that are processable by the system (40), whereby the context-enhanced database (49) is accessed in order to find matching segments for said segments;
- a module (55) for preparing the output (57) comprising the matching segments;
- a module (55) for enabling editing/correction of the output (57) in order to generate a final output (33) and for making the final output (33) available.

18. The system of claim 17, wherein the speech recognition system (51 – 56) processes words separately by

- identifying a matching word in the context-enhanced database (49) for each of the words, and
- adding the matching word to the output (57).

19. The system of claim 17 or 18, wherein the speech recognition system (51 – 56) analyzes the speech signals (50).

20. The system of claim 17 or 18, comprising another database (13) which is accessible in if no matching word is available in the context-enhanced database (49).

21. The system of claim 17 or 18, comprising a module that derives the input (22) from an application program.

22. The system of claim 17, wherein the input (41) is received or derived from one or more of:

- a history of E-mails,
- a document that is currently on a screen of the computer system,
- a chain of related documents,
- linked documents,
- a folder or directory,
- attachments that are received with an E-mail,
- a spread sheet,
- a cache memory of the computer system,
- a history information recorded by a web browser,
- a knowledge management system,
- an incoming message,
- a received facsimile, and
- from a result of a database search.

23. The system of claim 17 or 18, wherein the context-enhanced database (49) defines a context in light of which the voice-generated output is generated.

24. The system of claim 17 or 18, wherein the voice-generated output is a physical output.

25. The system of claim 17 or 18, comprising a memory (58) for storing the voice-generated output.

26. The system of claim 17 or 18, comprising a module that enables the editing/correction of the output.

27. The system of claim 17 or 18, comprising a database (61; 76, 77; 86) from which the context-enhanced database is derived.

28. The system of claim 17 or 18, comprising a synonym lexicon (61) that is linked when used.

29. The system of claim 17 or 18, comprising a meaning variants database (76) that is linked when used.

30. The system of claim 17or 18, wherein the module for automatically creating a context-enhanced database (49) is a pre-processing module.

31. The system of claim 17 or 18, comprising a meaning extraction system.

32. A computer program product comprising a computer readable medium having embodied therein computer program code which, when loaded in a computer system configures the computer system to perform the steps of:

- receiving an input (22);
- creating a context-enhanced database (21) by using information derived from the input (22);
- performing a speech recognition task to convert a speech signal (23) into an output (30) comprising computer processable segments, whereby the context-enhanced database (21) is accessed in order to improve the speech recognition rate;
- enabling editing of the output (30) in order to generate a final output (33); and
- making the final output (33) available.

33. A computer program element comprising computer program code which, when loaded in a computer system, configures the computer to perform a method for providing a voice-generated output as claimed in any of claims 1 to 16.

CH9-2000-0004

ABSTRACT

Method or system producing a voice-generated output (33). A context-enhanced database (21) is generated from a system input (22. The voice-generated output (30) is generated from a speech signal (23) by performing a speech recognition task to convert the speech signal (23) into computer processable segments. During this speech recognition task the context-enhanced database (21) is accessed in order to improve the speech recognition rate, i.e., to interpret the speech signal in light of words given in the context-enhanced database (21). A user is enabled to edit/correct the output (39) in order to generate a final output (33) which is then made available.

(Fig. 2)

CH9-2000-0004

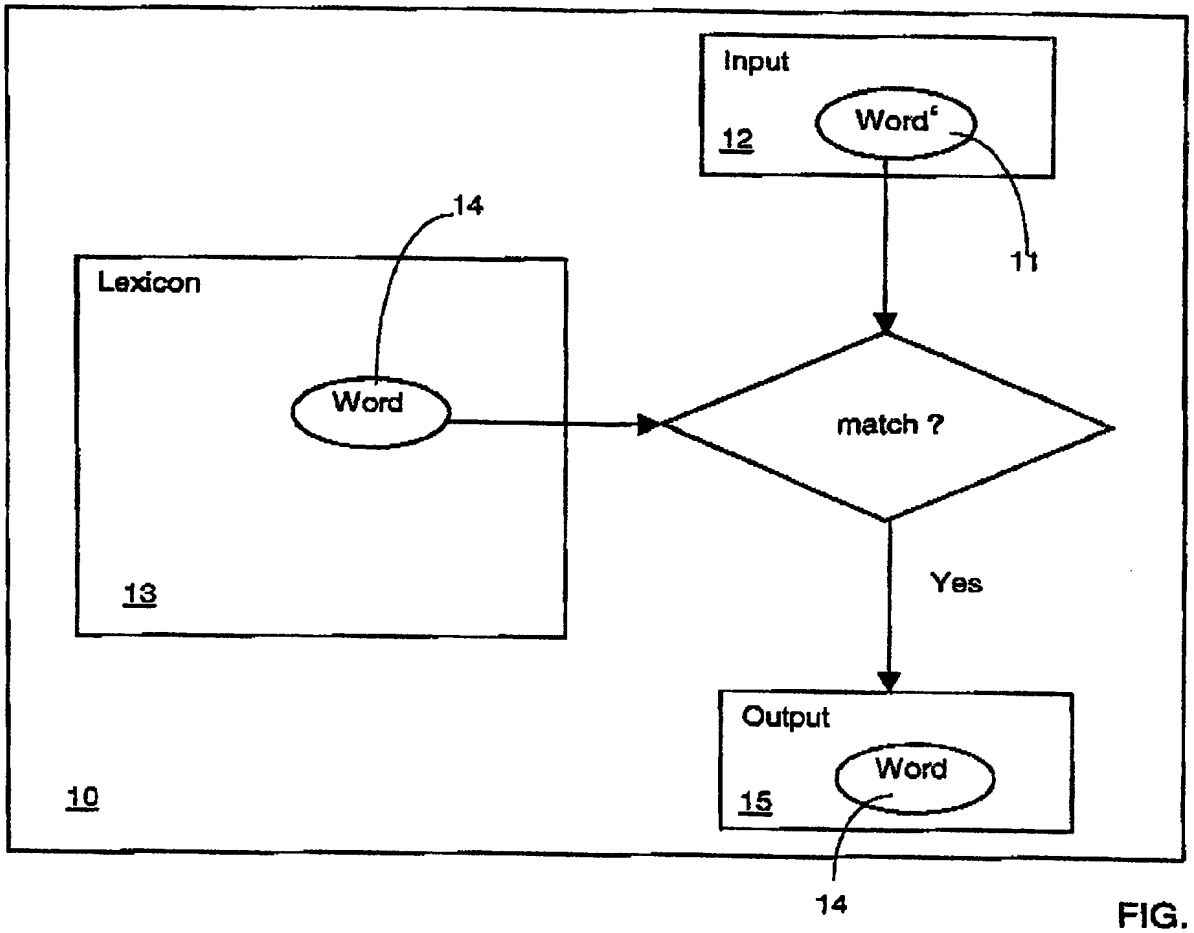

FIG. 1

CH9-2000-0004

2/5

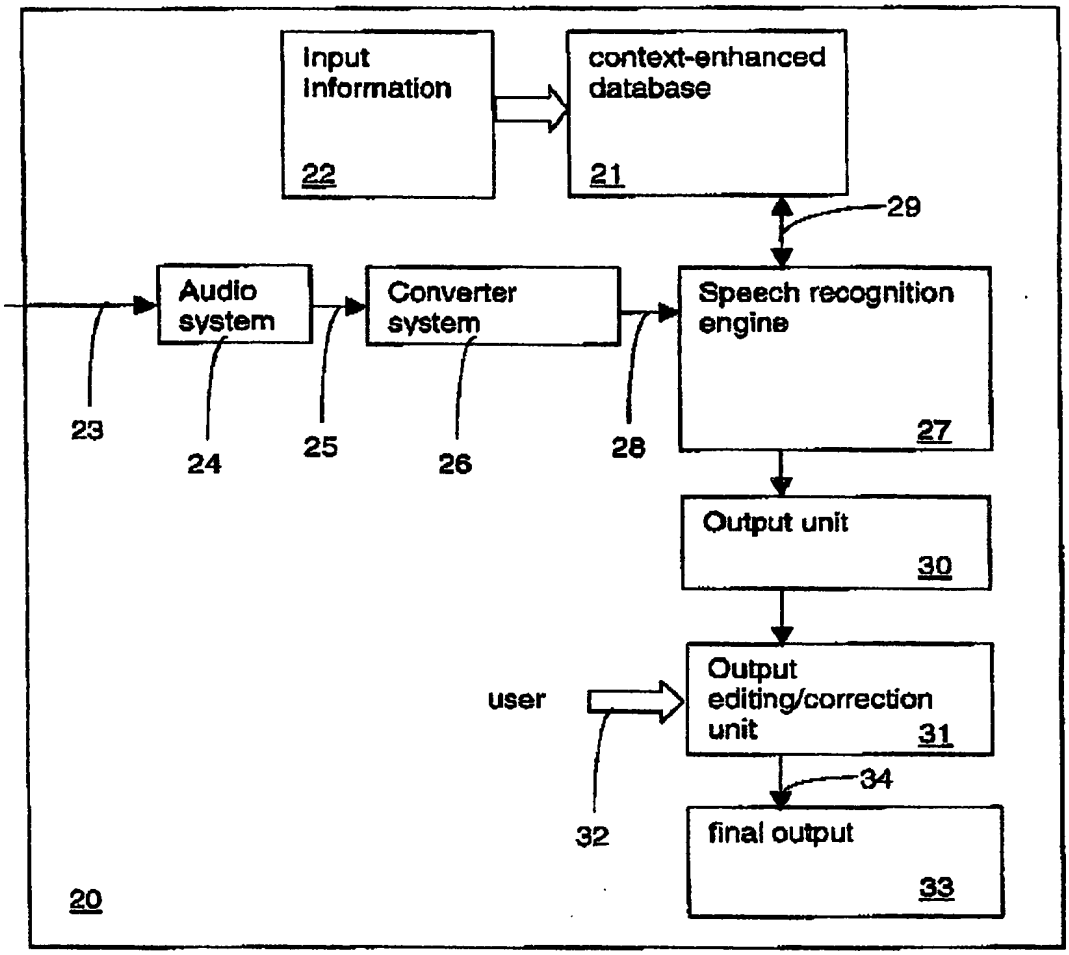

FIG. 2

CH9-2000-0004

3/5

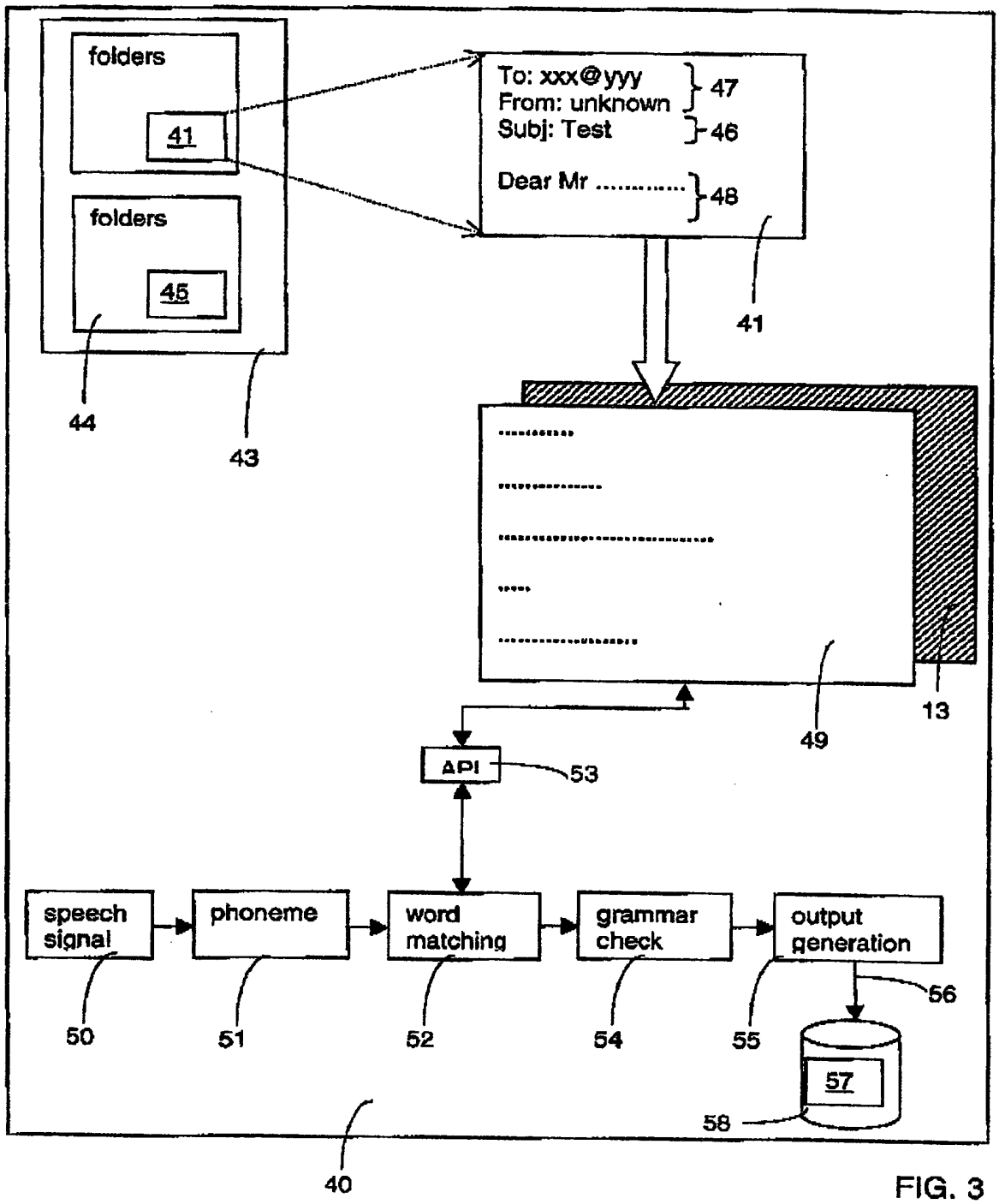

FIG. 3

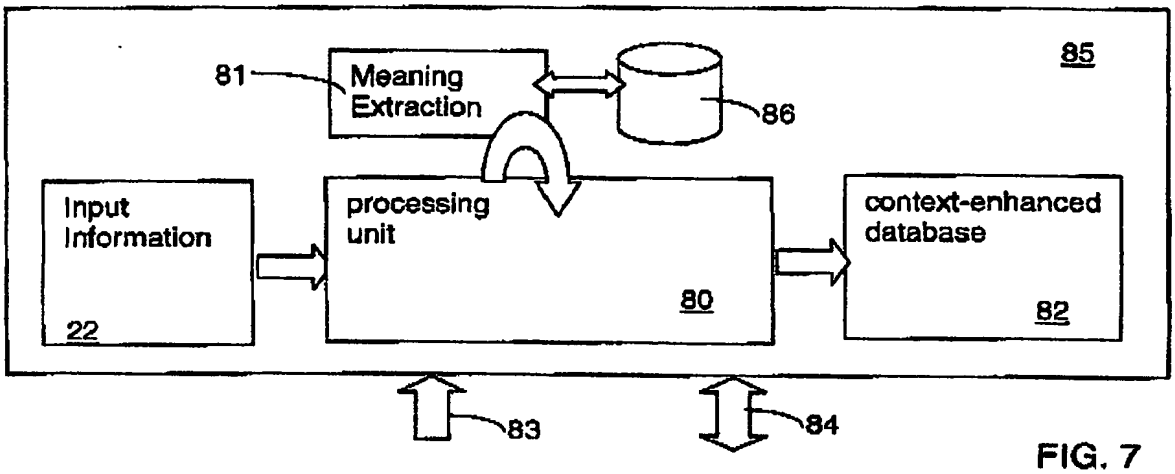
81
Meaning Extraction
86
85
Input Information
22
processing unit
80
context-enhanced database
82
83
84

FIG. 7

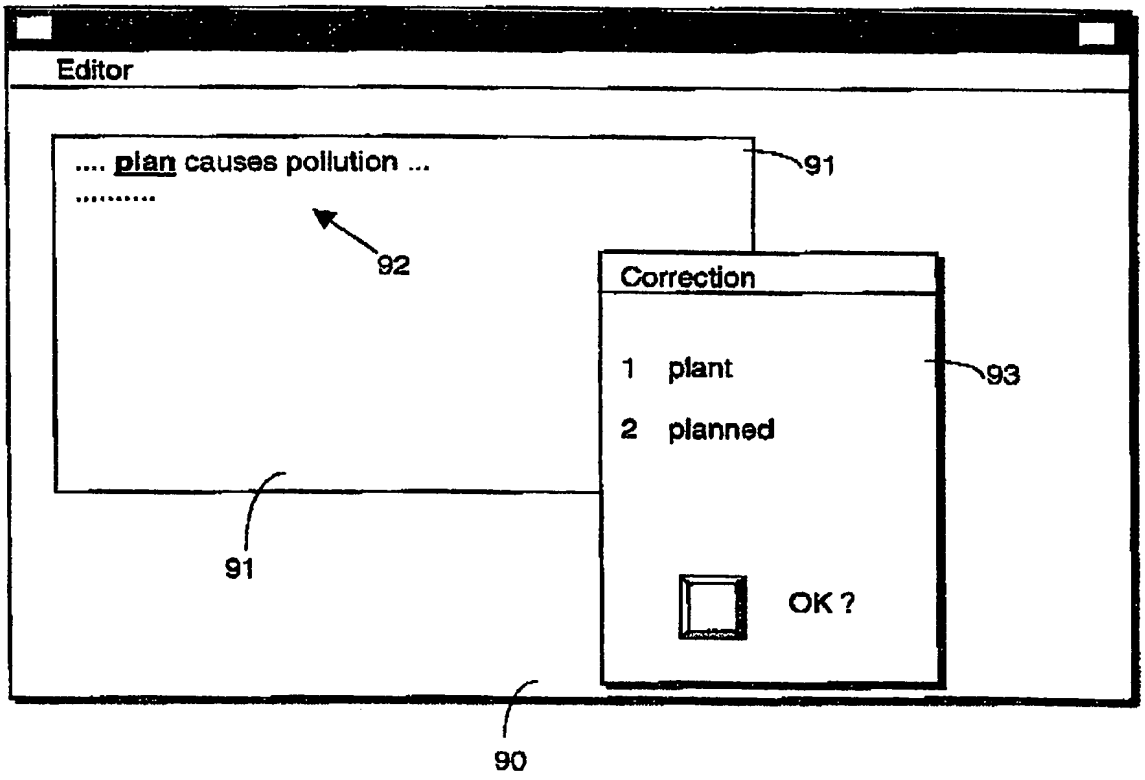
Editor
.... plan causes pollution ...
.....
91
92
Correction
1 plant
2 planned
93
91
OK ?
90

FIG. 8

CH9-2000-0004

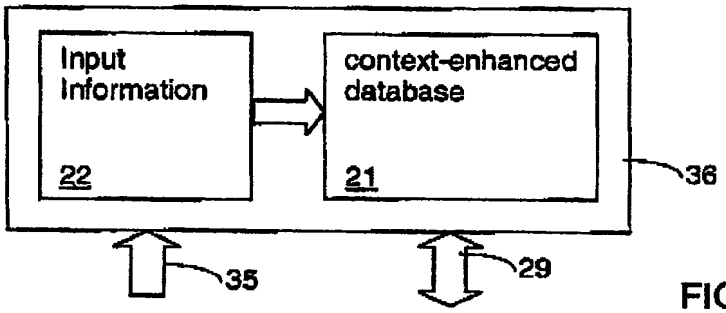

FIG. 4

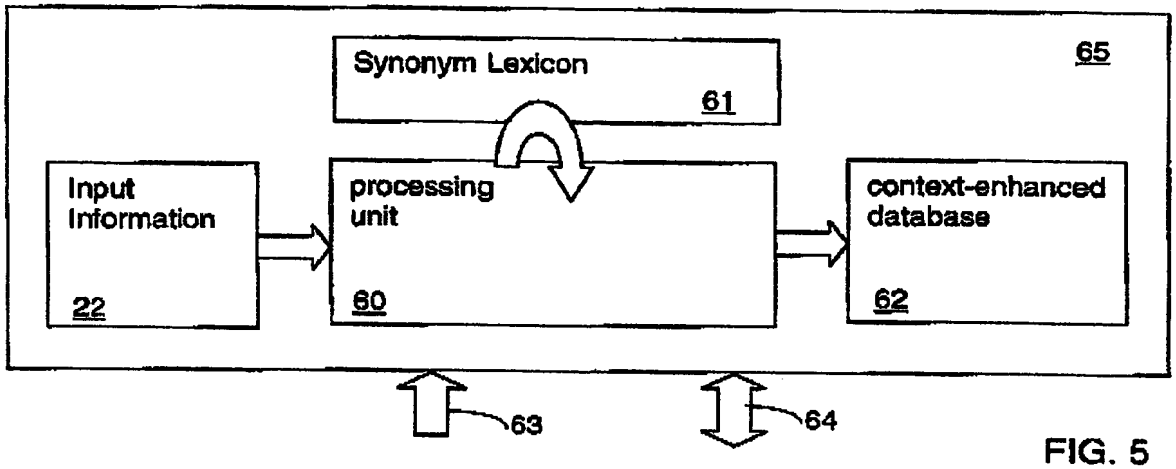

FIG. 5

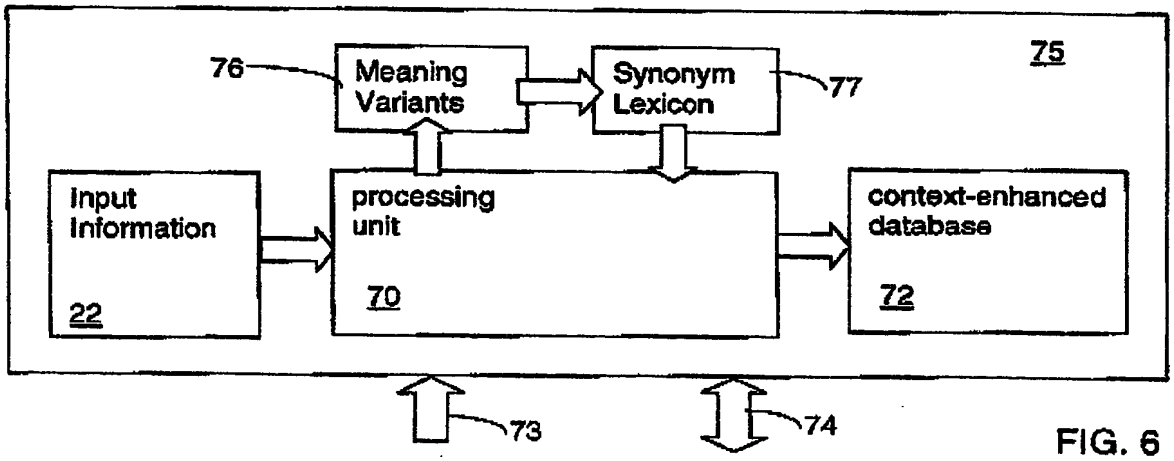

FIG. 6